**The Report Committee for Shijing Zhong**
**Certifies that this is the approved version of the following Report:**

# A Review on Constrained Recurrent Sparse Auto-Encoder

**APPROVED BY**
**SUPERVISING COMMITTEE:**

Chandrajit Bajaj, Supervisor

Clint Dawson

# A Review on Constrained Recurrent Sparse Auto-Encoder

**by**

**Shijing Zhong**

**Report**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Master of Science in Computational Science, Engineering, and Mathematics**

**The University of Texas at Austin**

**May 2020**

# Abstract

## A Review on Constrained Recurrent Sparse Auto-Encoder

Shijing Zhong, M.S.C.S.E.M

The University of Texas at Austin, 2020

Supervisor: Chandrajit Bajaj

Sparse Dictionary Learning generates a sparse representation for images and signals along with a generalized learned dictionary. We examine closely to the constrained recurrent sparse auto-encoder (CRsAE) on its Encoder-Decoder plus recurrent architecture and experimenting CRsAE's position in the classical dictionary learning problem. We further extend the visualizations, experiments, and metrics to evaluate the model in the context of both VAE and Dictionary Learning.

Table of Contents

# List of Figures

# INTRODUCTION

Sparse Dictionary Learning is a method for decomposing the image into a significant representations from the data such as signals and images. The representation is usually a vector of vector $x$ that corresponds to the image with a dictionary matrix as reference. The reconstruction of image involves a linear combination of significant patches(atoms) in the dictionary matrix, and the data can be encoded into the representation for processing. The representation becomes meaningful when its l0 norm is controlled, which is equivalent to being sparse. Therefore, the sparse dictionary learning is essentially a minimization problem $\min_x ||x|| \; s.t. \; y - Dx < \epsilon$. Traditional methods to solve CDL are K-SVD(Michal Aharon and Bruckstein, 2006), Method of optimal di rections(MOD)(Engan et al., 1999) and ISTA (Beck and Teboulle, 2009) algorithm that try to minimize the $l1$ norm to approximate the sparsity.

Improving from the traditional sparse coding learning, the convolutional sparse representations(Zeiler et al., 2010) change the reconstruction process of atoms in dictionary $D$. Given image $y$, traditional sparse dictionary approximates $y = Dx$. However, $D$ could be substantially large and redundant, and convolutional sparse representations suggest an approximation with $y = \sum_n d_n * x_n$ that can avoid this problem and generalize each atom. This representation leads to many design of convolutional dictionary learning(CDL) algorithm that involves similar to standard dictionary learning. A CDL algorithm usually consists of solving a convolutional sparse

coding(CSC) problem, updating the dictionary and the coupling mechanism in between these three parts(Garcia-Cardona and Wohlberg, 2018). The leading algorithm for optimizing sparse coding is Alternating Direction Method of Multipliers(ADMM)(Eckstein, 2012), but despite of a detailed convex analysis and computational advantages, it requires to fit the entire problem dataset into the GPU memory; without a well-managed distributed system, ADMM cannot be operated in presence of a large dimensional dataset. An more accessible algorithm to solve the CSC problem is using ISTA(Daubechies et al., 2003) based algorithm. Learned ISTA(LISTA)(Sreter and Giryes, 2018) suggests a recurrent network that approximate each ISTA iteration, and the CRsAE model we experiment uses modified ISTA to obtain sparse coding. The second part of CDL dictionary update can also apply with FISTA proposed in (Wohlberg, 2016). The iteration steps would involve updating $y^{(i)}$ and $d^{(i)}$, but it lacks the communication to the CSC step which leads to a complicated coupling mechanism(GarciaCardona and Wohlberg, 2018).

The model CRsAE inspired by Expectation Maximization algorithm then combines CSC step and dictionary update as an auto-encoder and decoder network.(Tolooshams et al., 2019) The architecture of the model uses a modified ISTA similar to LISTA(Sreter and Giryes, 2018) in encoding the sparse code, but improves on the coupling mechanism between the encoder and decoder with defined prior and likelihood function on regularization term. We examine CRsAE structure with a more straightforward insight and extends further experiments on its potential to bridge the connection to Variational Auto-Encoder in the following chapters.

# REVIEW OF CRSAE

## 1. Encoder-Decoder & CDL

In the classical dictionary learning, the optimization problem is equation (1). Since $||x||_0$ is a non-convex optimization problem, transforming into $l1$-norm relaxed this issue.

$$\text{argmin}_{x,\{d^i\}_{i=1}^N} \left\|x^i\right\|_1 s.t.\ y - Dx < v \tag{1}$$

$$\text{argmin}_{x,\{d^i\}_{i=1}^N} \frac{1}{2}\|y - Dx\|_2^2 - \lambda\left\|x^i\right\|_1 + v \tag{2}$$

CDL generalizes the dictionary from 2 dimensional filters into a set of filters for linear superstitions. The reconstruction of sparse coding is no longer combining patches into corresponds location, but linearly applying each atom into a convoluted image, and each atom $di$ is not restricted to size and have the freedom to be expanded even to the size of the whole image. The optimization problem has the following form

$$\text{argmin}_{\{x_i\}_{i=1}^N,\{d^i\}_{i=1}^n} \frac{1}{2}\left\|y - \sum_i d_i\, x_i\right\|_2^2 - \lambda\|x_i\|_1 + v \tag{3}$$

$y \in R^M$ , $x \in R^N$ , and d $d_i \in R^{MN}$ , because of the shifting from convolution superstition, and dimensions of atoms $d^i$ is not restricted to 2, and the size of $d^i$ also

doesn't need to necessarily match $y$. Because of notation, $d^i$ is zero-padded to the same size of $y$, but in implementation each $d_i x_i$ can simply perform a convolutional operator in figure 1. The smaller patches now can be convoluted into one combined output with the same dimensions of input image.

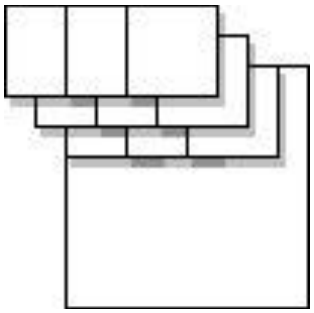

Figure 1: Convolutional operator on reconstruction, smaller patch is $x_i d_i$

The general optimization from (1) can always be generalized to two steps, one is convolutional sparse coding(CSC) update, and the second is convolutional dictionary update(Garcia-Cardona and Wohlberg, 2018). Compares to Variational Auto-Encoder(Kingma and Welling, 2013), which also performs a compression-decompression liked image processing, the first part of solving a CSC parallels to the encoder, and the convolutional dictionary update parallels to the back-propagation of decoder and encoder.

### 1.1 CSC IN CDL

In the setting of classical CDL, the CSC step's goal is to estimate a sparse representation coordinate of $X$ with the convolutional dictionary given. The problem can be rephrased into the following

$$\text{argmin}_{\{x_i\}_{i=1}^N} \frac{1}{2} \|y - K(d_i, x_i)\|_2^2 - \lambda \|x^i\|_1 \quad (4)$$

The first term is the error between original image signal and the convoluted reconstruction of the sparse code. Operator $K(d_i, x_i)$ represents the convolution reconstruction of sparse code and the dictionary visualized in figure 1. Normally after adding a zero padding for $d_i x_i$, the output can be a stacking of $d_i x_i$, but in actual implementation, the output can be modified in place according to their convoluted center.

While CSC parallels to the encoder of VAE, the Sparse Code search the layer doesn't involve with a multi-layer encoder with different parameters in different layers, and the dictionary $D$ is given and unless a different structure of dictionary is proposed (Kingma and Welling, 2013). Encoding of a regular CSC problem only aims to optimize equation (4) with an optimizer algorithm such as ADMM(Eckstein, 2012) or ISTA(Daubechies et al., 2003).

### 1.2 BACK PROPAGATION OF DICTIONARY UPDATE

After obtaining the sparse encoding of the dictionary, the dictionary also needs to update corresponds to in the dictionary learning step, the optimization problem is rephrased into

$$\mathrm{argmin}_{\{d_i\}_{i=1}^{MN}} \frac{1}{2} \|y - \mathrm{K}(\mathrm{d_i}, x_i)\|_2^2 s.t.\ \|d_m\|_2 = 1 \qquad (5)$$

In actual implementation, dimension of $d_i$ would be restricted to pre-defined dictionary size, which means the choice of choosing the numbers of atoms depends on tuning. The 2-norm restraint over dictionary will be reflected in a normalization over $D$ in each iteration. Dictionary Update can be solved with an optimizer algorithm, but a back propagation of

the loss function which updates each atom $d_i$(Tolooshams et al., 2018) serves the same purpose. Because of the similarity of loss function (4) and (5), gradient of the atom $d_i$ obtained from optimizer can be updated with back propagation, and therefore the model can save a duplicated optimizer that only updates the atom $d_i$.

# 2. Architecture Examination of CRsAE

CRsAE is a recurrent two-step neural network that mimics the design of auto-encoder and decoder. The model can be divided into two part, the encoder part is essentially a CSC step that performs FISTA to estimate a sparse code given a convoluted dictionary $D$. Similar to the work of the Learned Convolutional Sparse Coding(LSC) model (Sreter and Giryes, 2018), the encoder has a recurrent structure that compresses ISTA into one forward propagation. The decoder is essentially the dictionary update of CDL, which is handled by the backpropagation of the network. However, compared to LSC, CRsAE proposed a coupling mechanism that combine encoder and decoder with the "constrains" that derived from modifying the regularization parameter $\lambda$(Tolooshams et al., 2019).

## 2.1 Alternating Step Between CSC and Dictionary Update

The recurrent neural network underlies in the CSC step in fact is equivalent to the ISTA(Daubechies et al., 2003) algorithm. In each iteration, update of x t corresponds to the optimization problem in (4) can be expressed as follows

$$x_k = \mathcal{P}_{\mathcal{L}}\left(x_{k-1} + \frac{1}{L}D^T(y - Dx_{k-1})\right) \tag{6}$$

FISTA has a faster convergence rate of $O(\frac{1}{t}^2)$ (Beck and Teboulle, 2009) over ISTA's convergence rate of $O(\frac{1}{t})$, because of the aggregated step size of $\frac{1+\sqrt{1+4t_{k-1}^2}}{2}$.

FISTA is nested inside the encoder and that causes the recurrent structure of the network, even though essentially is a coordinate descent combines within. Therefore, the number of iteration T dictates the sparsity of the code given the dictionary. Classical

CDL training requires an alternation between encoder and decoder to fit both sparsity of representation and improve dictionary over the sample(Garcia-Cardona and Wohlberg, 2018). CRsAE failed to mention that in training one pass for each sample cannot guaranteed the sparsity of dictionary, it requires training of multiple iteration over the same sample. Our hypothesis is that altering the encoder and decoder with a certain number of $T$ allows the model to obtain the new sparse representation given an updated dictionary.

### 2.2 CONNECTION TO VAE

VAE uses a constructed prior and posterior that optimized the Evidence Lower Bound(ELBO)(Kingma and Welling, 2013) due to the intractability of the marginal likelihood function of $p_\theta(z|x)$. The objective function of CDL is (3), but the similar technique of prior construction allows the network to back propagates. In a recent work of VAE with mixture of posteriors(Takahashi et al., 2018), it reconstructs the prior of latent variable $p(z)$ to reduce the KL-divergence in the objective function. The implementation of CRsAE left prior of dictionary matrix $D$ as flat, and leave the constraint of $H$ to a normalization in each iteration(Tolooshams et al., 2018). A further work can be done by constraining $\|D\|_2$with the prior of $P(D)$ instead of a normalization.

### 2.3 PRIOR DISTRIBUTION OF PARAMETERS

The CSC step and the update of convolutional dictionary is often treated as a separate problem to tackle. The review of Convolutional Dictionary Learning CDL(Garcia-Cardona and Wohlberg, 2018) mentioned dictionary update algorithm in its section III such as ADMM Consensus(Sorel and Sroubek, 2016) and FISTA(Beck and Teboulle, 2009), but they failed to update both parameters in dictionary and sparse code in one propagation. CRsAE derives prior distributions for dictionary, regularization parameter $\lambda$ and noise together(Tolooshams et al., 2019). From the optimization expression (3), a prior distribution for each parameter and likelihood function can be derived. Input image is approximated with a multivariate distribution $\mathrm{y|x, D, \sigma\hat{}2} \sim \mathcal{N}(\mathrm{Dx}, \sigma\hat{}2)$; likelihood function for $p(x|\lambda)$ can be approximated with Laplace probability density functions, and a joint prior $P(x, D, \lambda)$ of all parameters together assumes each independence.

$$log\big(P(y, x, D, \lambda)\big) = log\big(P(y|x, D, \lambda)P(x, D, \lambda)\big) \tag{7}$$

$$= l - \frac{1}{2\sigma^2}\big||y - Dx|\big|_2^2 - \lambda\big||x|\big|_1 + ClogP(\lambda)P(D) \tag{8}$$

With expansion of the above probability, CRsAE model can use back propagation through the network to obtain additional update of $\lambda$. Dictionary $D$ can be updated through back propagation of the encoder recurrent FISTA. This coupling allows model to compress both CSC and Dictionary learning into one network instead of two. In fact theCompares to Learned Convolutional Sparse Coding (Sreter and Giryes, 2018), the

addition of prior creates new potential on constructing different prior to further approximate the optimization constraint.

## 3. Experiments

The original paper has conducted various signal denoising test(Tolooshams et al., 2019), but have not addressed the sparsity of the encoded representation. Sparsity of the encoding signal x can represent the importance of $\lambda$ learning, and effectiveness of categorize the significant feature of the input. The encoding iteration of FISTA is equivalent to the depth of the recurrent network, and from the report it suggests a significant improvement in performance of denoising, to test this hypothesis, we visualize the sparse representation's latent space as well as comparing PSNR with different depth. Another insight we explore is the effect of alternation between CSC step and dictionary update step. CRsAE treats every training sample as one alternation as backward propagation by ignoring the potential change of sparse representation after dictionary's update.

Our experiments are conducted on two data sets, MNIST(LeCun and Cortes, 2010) and Visual Object Classes(VOC) in 2012(Everingham et al., 2015). In the MNIST data set, we visualizes the sparse coding in the latent space using the popular manifold learning based algorithm t-Distributed Stochastic Neighbor Embedding (tSNE)(van der Maaten and Hinton, 2008). We trained CRsAE for denoising task on this task, which we crop a $250 \times 250$ windows for each image with 20 standard deviation noise for training and evaluation. Dictionary D has 12 separated $15 \times 15$ patches with the convolution of 6 stride. Optimization is done with a cyclic step scheduler over ADAM(Kingma and Ba, 2014) in batch size of 10. The length of the encoder parameter $T$ is part of the comparison below, and in general testing we fixed it to 10. Alternation step remains to 1 if not mentioned.

### 3.1 LEARNING CONVERGENCE

Losses in training represents the stability of the model in training. Same parameters are used in the denoising subsection, and in this section, we focus on visualizing the convergence. We use a supervised loss with the ground true image in training of denoising task for VOC. The parameter we mainly compare to is the depth of the encoder T and the alternating encoder-decoder step. We find that a deeper recurrent encoder leads to a more stable training loss. It lines up with our observation that the encoder FISTA corresponds to the CSC step which is essential for CDL.

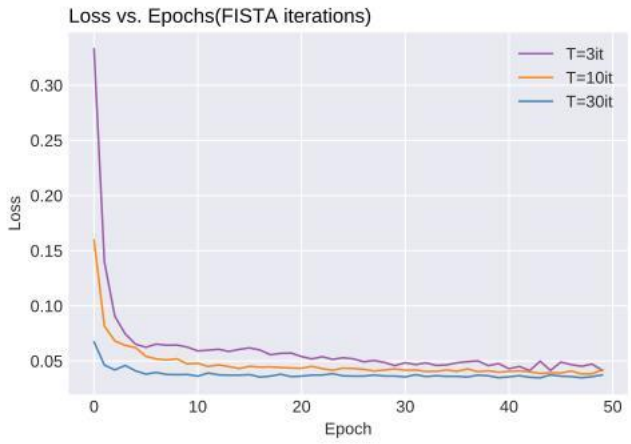


Figure 2: Convergence plot of different depths of recurrent FISTA

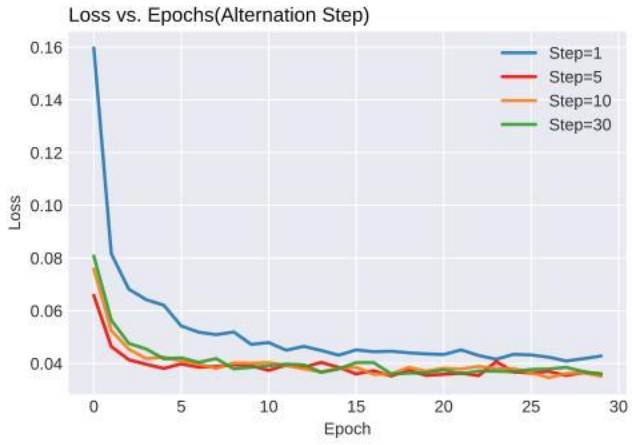


Figure 3: Convergence plot of different alternating CSC and dictionary learning step

CDL problem often involves an alternation between CSC step and dictionary learning step(Garcia-Cardona and Wohlberg, 2018). The decoder $D^T x$ performs a convolutional operator (Fig 1.) to reconstruct the original image. In each iteration, we added fixed alternation step between algorithm 1 and the decoder. Performing additional alternation steps clearly stabilizes the loss of each step. This observation indicates that CRsAE can be stabilized with additional sparse coding step after decoder update $D$.

### 3.2 DENOISING

CRsAE was demonstrated to successfully denoise images from VOC, and samples of denoised images are listed in Figure 8. The Peak Signal-to-Noise Ratio(PSNR) is the evaluation criterion of image denoising. In each iteration, PSNR obtained from the noised image and denoised image reflects the performance of CRsAE. Figure 4 shows a comparison of different $T$ length of encoder's performance, and it reflects that convergence of encoder can substantially improve denoising as well. Similarly, alternation step size greater than 1 leads to a faster convergence along with performance of denoising.

Therefore, the classical sparse coding search and dictionary update still explains the encoder-decoder model and improve its performance.

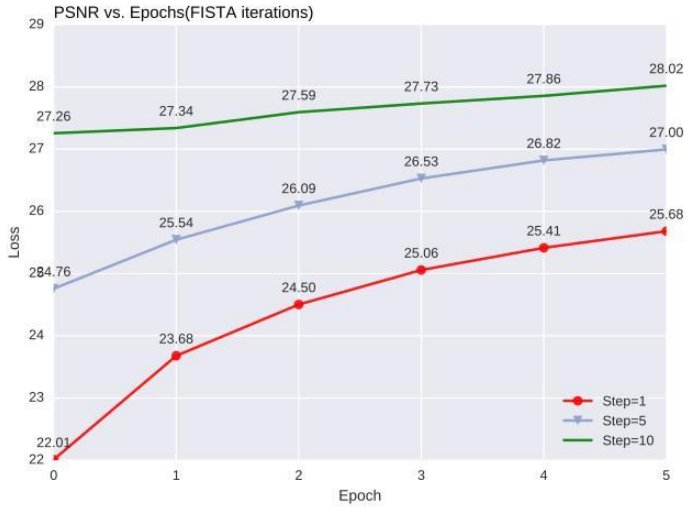


Figure 4: PSNR of different depths of recurrent FISTA

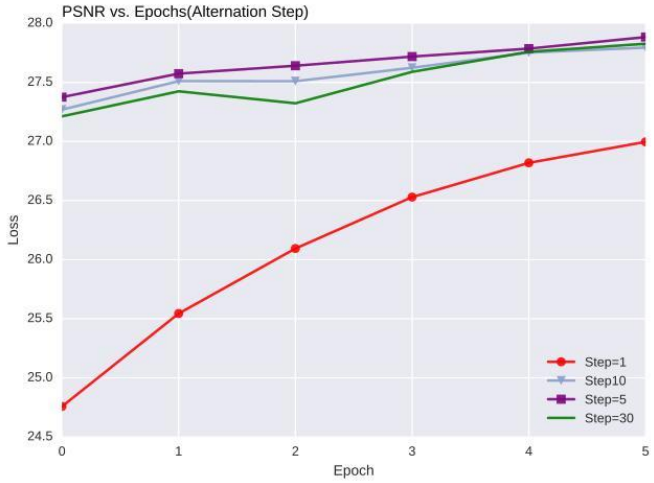


Figure 5: PSNR of different alternating CSC and dictionary learning step

### 3.3 LATENT SPACE OF SPARSE REPRESENTATION

In the context of analyzing embedding of the VAE model, examining the embedding distribution in the latent space distribution is usually an option. The

convolutional sparse representation has lower the dimension of the input dimension, but still has a high-dimension that is difficult to visualize, and many proposed learning algorithm can serve as dimension reduction techniques such as PCA.(Sehgal et al., 2014) We are using a popular manifold learning based algorithm tSNE to visualize the sparse representation.(van der Maaten and Hinton, 2008) The algorithm accumulates sparse representation in to cluster by minimizing the KL divergence of the cluster distribution and Euclidean distance.

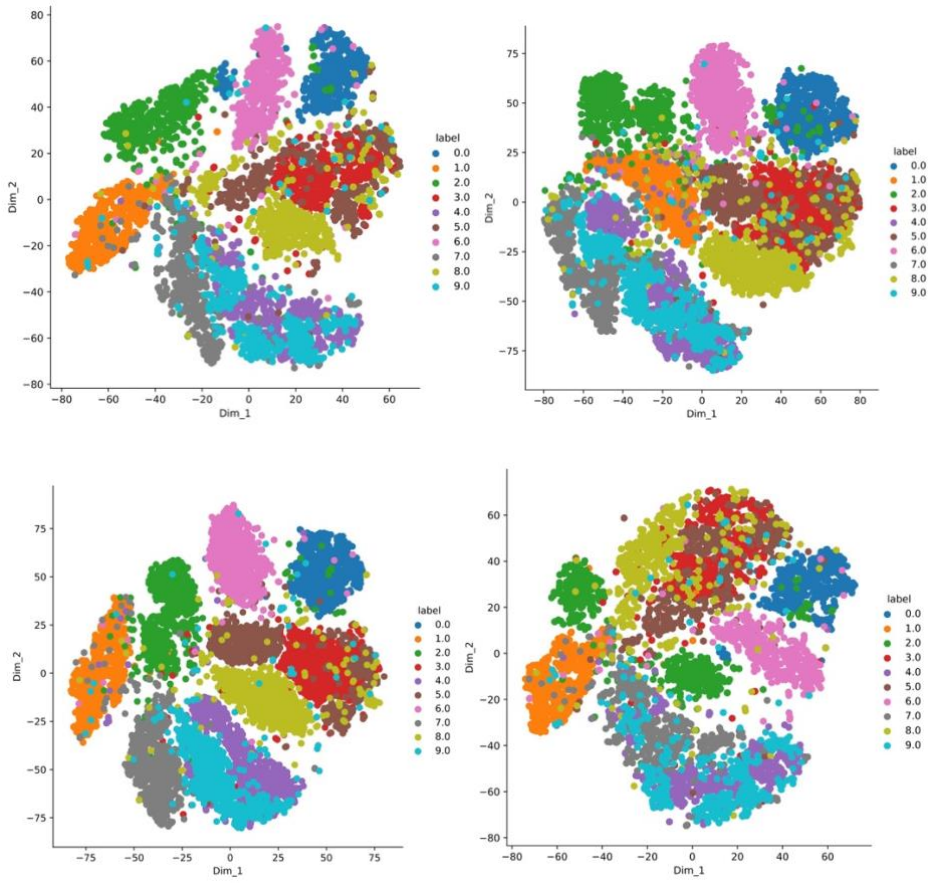


Figure 6: Cluster of label images being separated within first iteration, but it separates after one epoch, started from left to right, top and down

Figure 6 is the tSNE clustering of digits from 0-9 in MNIST data set with dimension of $28 \times 28$. The dictionary size for MNIST is $64 \times 15 \times 15$ with stride of 6, and sparse coding x has a dimension of $36 \times 64 \times 6 \times 6$(Tolooshams et al., 2018). Figure 6 only shows several iterations change within one epoch. CRsAE can separate each cluster in the few iterations because each dictionary filters haven't assimilated due to its convolution nature, and clusters get entangled a few epochs. Despite of a prevalent usage of tSNE to visualize, tSNE is not suitable for sparse coding representation.

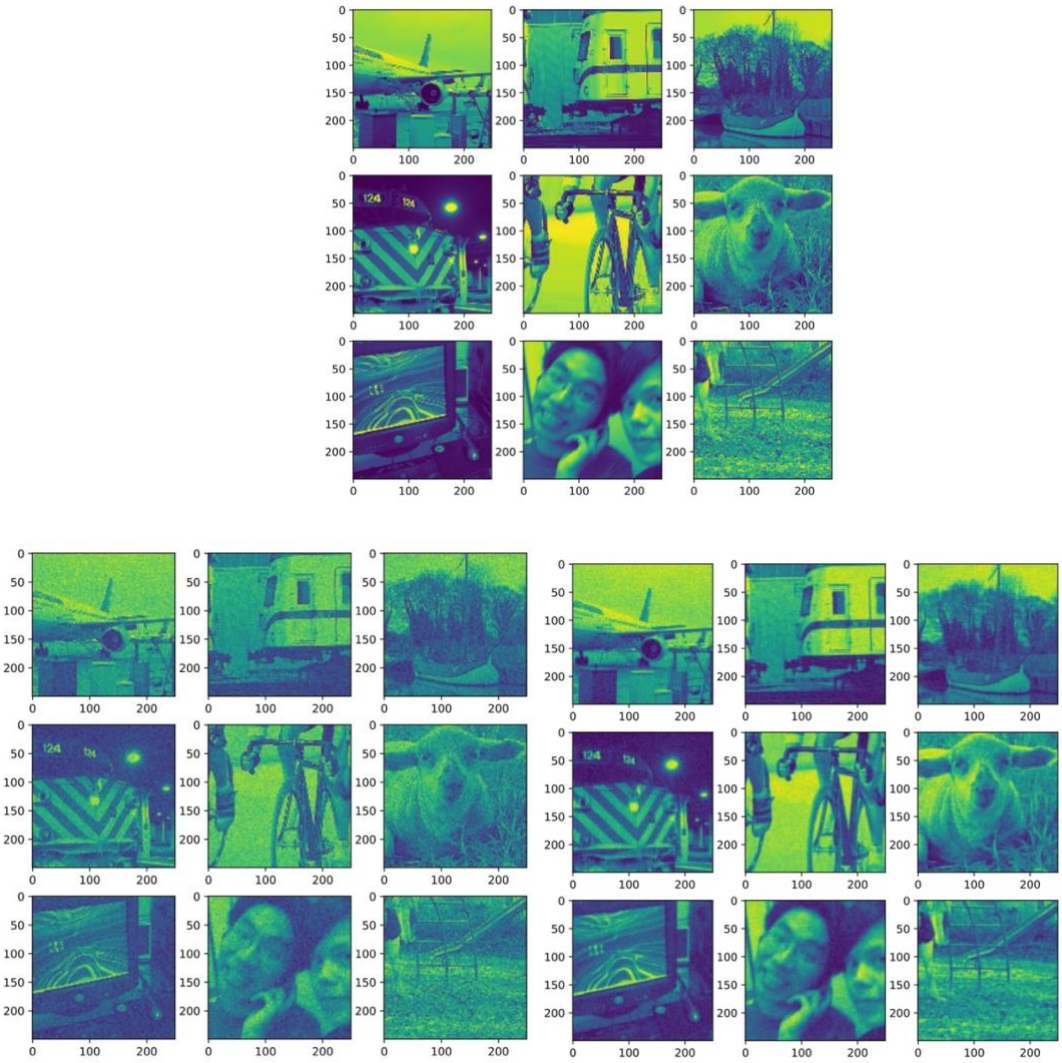

Figure 7: Convolutional operator on reconstruction, smaller patch is $x_i d_i$

Top image is original samples, bottom left is noised version and bottom right is denoised image

# CONCLUSION

This review summarizes the novel CRsAE model that adapts the recurrent neural network and constrained prior distribution on hyperparameters. Comparing the model with Learned Convolutional Sparse model(Sreter and Giryes, 2018), CRsAE approximated $P(x, D, \lambda)$ and $P(x \mid \lambda)$ with multivariate Gaussian and Laplace distributions to achieve back propagation to dictionary update. Even with the structural similarity towards autoencoders, CRsAE remains the nature of solving a classical CDL problem. We assume that the recurrent structure takes the form of CSC update, and provided experimental results on the importance of iteration depth. In the other experiments, we obtained stable improvement on increasing the alternating step between sparse coding and dictionary update. The last experiment of visualizing latent space of sparse coding indicates a different ideology between VAE and sparse coding. In future work, a detailed comparison between VAE and CRsAE can explain these discrepancies.

# REFERENCES


Amir Beck and Marc Teboulle. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences 2(1):183–202. https://doi.org/10.1137/080716542.

Ingrid Daubechies, Michel Defrise, and Christine De Mol. 2003. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.

Jonathan Eckstein. 2012. Augmented lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results.

Ingrid Daubechies, Michel Defrise, and Christine De Mol. 2003. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.

K. Engan, S. O. Aase, and J. Hakon Husoy. 1999. Method of optimal directions for frame design 5:2443–2446 vol.5.

M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. 2015. The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision 111(1):98–136.

Cristina Garcia-Cardona and Brendt Wohlberg. 2018. Convolutional dictionary learning: A comparative review and new algorithms. IEEE Transactions on Computational Imaging 4(3):366–381. https://doi.org/10.1109/tci.2018.2840334.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization.

Diederik P Kingma and Max Welling. 2013.Auto-encoding variational bayes.

Yann LeCun and Corinna Cortes. MNIST handwritten digit http://yann.lecun.com/exdb/mnist/.

Michael Elad Michal Aharon and Alfred Bruckstein. 2006. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Computational Imaging http://dx.doi.org/10.1109/TCI.2018.2840334.

S. Sehgal, H. Singh, M. Agarwal, V. Bhasker, and Shantanu. 2014. Data analysis using principal component analysis. In 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom). pages 45–48.

Hillel Sreter and Raja Giryes. 2018. Learned convolutional sparse coding. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) https://doi.org/10.1109/icassp.2018.8462313.

Bahareh Tolooshams, Sourav Dey, and Demba Ba. 2018. Scalable convolutional dictionary learning with constrained recurrent sparse auto-encoders.

Bahareh Tolooshams, Sourav Dey, and Demba Ba. 2019. Deep residual auto-encoders for expectation maximization-inspired dictionary learning.

Laurens van der Maaten and Geoffrey E. Hinton. 2008. Visualizing data using t-sne.

Michal Sorel and Filip Sroubek. 2016. Fast convolutional sparse coding using matrix inversion lemma. Digit. Signal Process. 55(C):44–51. https://doi.org/10.1016/j.dsp.2016.04.012.

B. Wohlberg. 2016. Efficient algorithms for convolutional sparse representations. IEEE Transactions on Image Processing 25(1):301–315.

M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus. 2010. Deconvolutional networks pages 2528–2535.